Central Cancer Registry Receive Batch File Use Case

Version 1.0

Prepared by: NPCR–MERP Central Cancer Registry Workgroup

NPCR–MERP Technical Development Team

Centers for Disease Control and Prevention

National Center for Chronic Disease Prevention and Health Promotion

Division of Cancer Prevention and Control

National Program of Cancer Registries

March 2, 2008

Table of Contents

General Information

1. Use Case ID

CCRUC 1.2

2. Use Case Name

Receive Batch File

3. Description

This use case describes the process for receiving a batch file from trusted data sources into the central cancer registry (CCR) database.

4. Actors

- Central cancer registry (CCR) software
- Data source software
- CCR staff

5. Definitions

- **Event Report:** The generic name used for a submission of data from a data source. It can be a cancer registry abstract or an electronic submission of an electronic health record (EHR) report such as a pathology or X-ray report.

- **Abstracted Event Report:** An event report that has been created by a registrar or data source personnel that includes information from multiple health records.

- **EHR Event Report:** An event report from an electronic health record. It may include data from multiple databases within a facility; however, no human evaluation or determination of data values is made. Additionally, no trained personnel have evaluated its relevance or reportability prior to submission.

- **Trusted Data Source:** A data source who meets the requirements for sending data electronically on a routine basis.

Receive Batch File

Note: Diagrams for this use case are in Appendix A and Appendix B.

1.0 Preconditions

A set of conditions that must be met before the activities described in the use case can begin.

1. The batch file is received electronically.

2. The batch file entering the CCR software is from a trusted data source.

2.0 Post Conditions

A set of conditions that must be met after the activities described in the use case have been completed.

The received batch file has been accepted to go forward.

3.0 Priority

Describes the importance and sequence of the use case in the overall activities of the cancer registry.

This is a high-priority use case and the workgroup decided it should be developed first. It follows the 1.1 Prepare and Transmit Event Report use case and precedes the 1.3 Validate Event Report use case.

4.0 Frequency of Use

Describes how often the activities in the use case take place.

The activities in this use case will take place each time a new or resubmitted batch file arrives from a trusted data source.

5.0 Normal Course of Events

Describes the specific steps taken to perform the activity in the use case.

Normal refers to the steps that are taken when everything goes according to routine procedures. Problems and exceptions are described in section 6, Alternative Course.

Business rules are statements that describe a decision that must be made and agreed to by those involved in the activity. In the context of this document, a business rule describes the decision that needs to be made, and in some circumstances provides a recommendation; in others, options for consideration and use.

Software requirements are statements that describe the functionality of the software that is required or recommended.

5.1 This use case begins when CCR software retrieves the batch file from the receiving data server.

5.2 CCR software loads the batch file from a trusted data source.[1] [BR01]

Note: A business rule will be added here that defines a trusted data source when the use case "Establish a Trusted Data Source" is finalized.

BR	Business Rule	Purpose	Remarks
01	The time interval for reporting to the CCR should be established for each data source.	To ensure timely reporting.	State regulations must be followed. The time interval is based on the number of reports and resources.

5.3 CCR software decrypts the batch file.

5.4 CCR software logs the batch file as received. [BR02, BR03]

BR	Business Rule	Purpose	Remarks
02	The batch file log should include recommended data items.	To ensure the ability to track and monitor batch submissions accurately.	See Appendix C for a list of data items to include in the batch file log.
03	A standard naming convention should be used for the batch files. See Appendix D.	To provide a national naming standard and track the files submitted.	The proposed format is DataSource.[ReportType].[SubmitType].DateOfTransmission.FileNumber

5.5 CCR software stores the batch file in a temporary workspace on the CCR computer system.

5.6 CCR software validates the record layout format for the batch file. [BR04, BR05, BR06]

BR	Business Rule	Purpose	Remarks
04	An acceptable version of the NAACCR Record Layout Version must be used for submitting cancer registry event report (abstracts).	To achieve uniformity and consistency.	The supporting documentation is: • NAACCR's *Data Exchange Standards and Records Description Vol.* I • NAACCR's *Data Standards and Data Dictionary Vol. II* Available at www.naaccr.org/index.asp?Col_SectionKey=7&Col_ContentID=122

---

[1] The guidelines for declaring a data source as a "trusted" data source are a separate use case.

BR	Business Rule	Purpose	Remarks
05	The NAACCR standard should be used for submitting pathology event reports electronically.	To achieve uniformity and consistency.	The supporting documentation is *Pathology Laboratory Electronic Reporting Version 2.0* (November 2005), available at www.naaccr.org/filesystem/pdf/Standards%20Volume%20V%20Final%20PDF%201-24-06.pdf
06	The CCR and the data source must determine an acceptable record layout format when no NAACCR standard exists.	To achieve uniformity and consistency.	Example: Physician and clinic records do not have a NAACCR transmission standard.

5.7 CCR software determines that the batch file is not a duplicate of a previous submission. [BR07]

BR	Business Rule	Purpose	Remarks
07	An electronic signature for the batch as a whole should be created and stored in the database. See Appendix E.	To prevent reprocessing of batches.	The electronic signature prevents a batch file from being processed more than once. Scenarios include: • A data source may submit the same batch multiple times. • The CCR may mistakenly try to process the same batch twice.

5.8 CCR software determines that the event report is not a duplicate of an existing event report in the CCR database. [BR08]

BR	Business Rule	Purpose	Remarks
08	CCR software should perform a deterministic record-by-record and data item-by-data item match on the following data items: • Last name • First name • Sex • Social Security number • Date of birth • Primary site • Laterality • Date of diagnosis • Morphology (histology / behavior) There may be a performance issue to check pathology reports data item-by-data item, so a subset of data items may be used.	To confirm that the batch file is a new submission.	

5.9 CCR software loads the batch file into the CCR database.

6.0 Alternative Course of Events

Numbering in this section refers to its associated step above in section 5, Normal Course of Events.

5.1a CCR software actively retrieves a batch file from a data source.

5.6a The batch file is not in a correct record layout format.

> **5.6a.1** CCR software rejects the batch file.

> **5.6a.2** CCR software notifies the data source (software) that batch file is rejected.

> **5.6a.3** CCR software updates the batch file log to indicate that the batch file is rejected.

> **5.6a.4** End of use case.

5.7a The batch file is an exact duplicate of a previously submitted batch file. [BR07]

> **5.7a.1** CCR software marks it as a duplicate.

> **5.7a.2** CCR software notifies the data source that the batch file has been submitted previously.

> **5.7a.3** CCR software updates the batch file log to indicate that the batch file is rejected.

> **5.7a.4** End of use case.

BR	Business Rule	Purpose	Remarks
07	An electronic signature for the batch as a whole should be created and stored in the database. See Appendix E.	To prevent reprocessing of batches.	The electronic signature prevents a batch file from being processed more than once. Scenarios include: • A data source may submit the same batch multiple times. • The CCR may mistakenly try to process the same batch twice.

5.8a The batch file contains exact duplicate event reports. [BR08]

5.8a.1 CCR software marks the event report as a duplicate and stores it.

BR	Business Rule	Purpose	Remarks
08	CCR software should perform a deterministic record-by-record and data item-by-data item match on the following data items: • Last name • First name • Sex • Social Security number • Date of birth • Primary site • Laterality • Date of diagnosis • Morphology (histology / behavior) There may be a performance issue to check pathology reports data item-by-data item, so a subset of data items may be used.	To confirm that the batch file is a new submission.	

7.0 Business Rules

A statement that describes a decision that must be made and agreed to by those involved in the activity. In the context of this document, a business rule describes the decision that needs to be made, and in some circumstances provides a recommendation; in others, options for consideration and use.

Business rules for this use case are presented under the step to which they apply.

BR	Business Rule	Purpose	Remarks
01	The time interval for reporting to the CCR should be established for each data source.	To ensure timely reporting.	State regulations must be followed. The time interval is based on the number of reports and resources.
02	The batch file log should include recommended data items.	To ensure the ability to track and monitor batch submissions accurately.	See Appendix C for a list of data items to include in the batch file log.
03	A standard naming convention should be used for the batch files. See Appendix D.	To provide a national naming standard and track the files submitted.	The proposed format is DataSource.[ReportType].[SubmitType]. DateOfTransmission.FileNumber
04	An acceptable version of the NAACCR Record Layout Version must be used for submitting cancer registry event report (abstracts).	To achieve uniformity and consistency.	The supporting documentation is: • NAACCR's *Data Exchange Standards and Records Description Vol.* I • NAACCR's *Data Standards and Data Dictionary Vol. II* Available at www.naaccr.org/index.asp?Col_SectionKey=7&Col_ContentID=122
05	The NAACCR standard should be used for submitting pathology event reports electronically.	To achieve uniformity and consistency.	The supporting documentation is *Pathology Laboratory Electronic Reporting Version 2.0* (November 2005), available at www.naaccr.org/filesystem/pdf/Standards%20Volume%20V%20Final%20PDF%201-24-06.pdf
06	The CCR and the data source must determine an acceptable record layout format when no NAACCR standard exists.	To achieve uniformity and consistency.	Example: Physician and clinic records do not have a NAACCR transmission standard.
07	An electronic signature for the batch as a whole should be created and stored in the database. See Appendix E.	To prevent reprocessing of batches.	The electronic signature prevents a batch file from being processed more than once. Scenarios include: • A data source may submit the same batch multiple times. • The CCR may mistakenly try to process the same batch twice.

BR	Business Rule	Purpose	Remarks
08	CCR software should perform a deterministic record-by-record and data item-by-data item match on the following data items: • Last name • First name • Sex • Social Security number • Date of birth • Primary site • Laterality • Date of diagnosis • Morphology (histology / behavior) There may be a performance issue to check pathology reports data item-by-data item, so a subset of data items may be used.	To confirm that the batch file is a new submission.	

8.0 Exceptions

None.

9.0 Includes

None.

10.0 Special Requirements

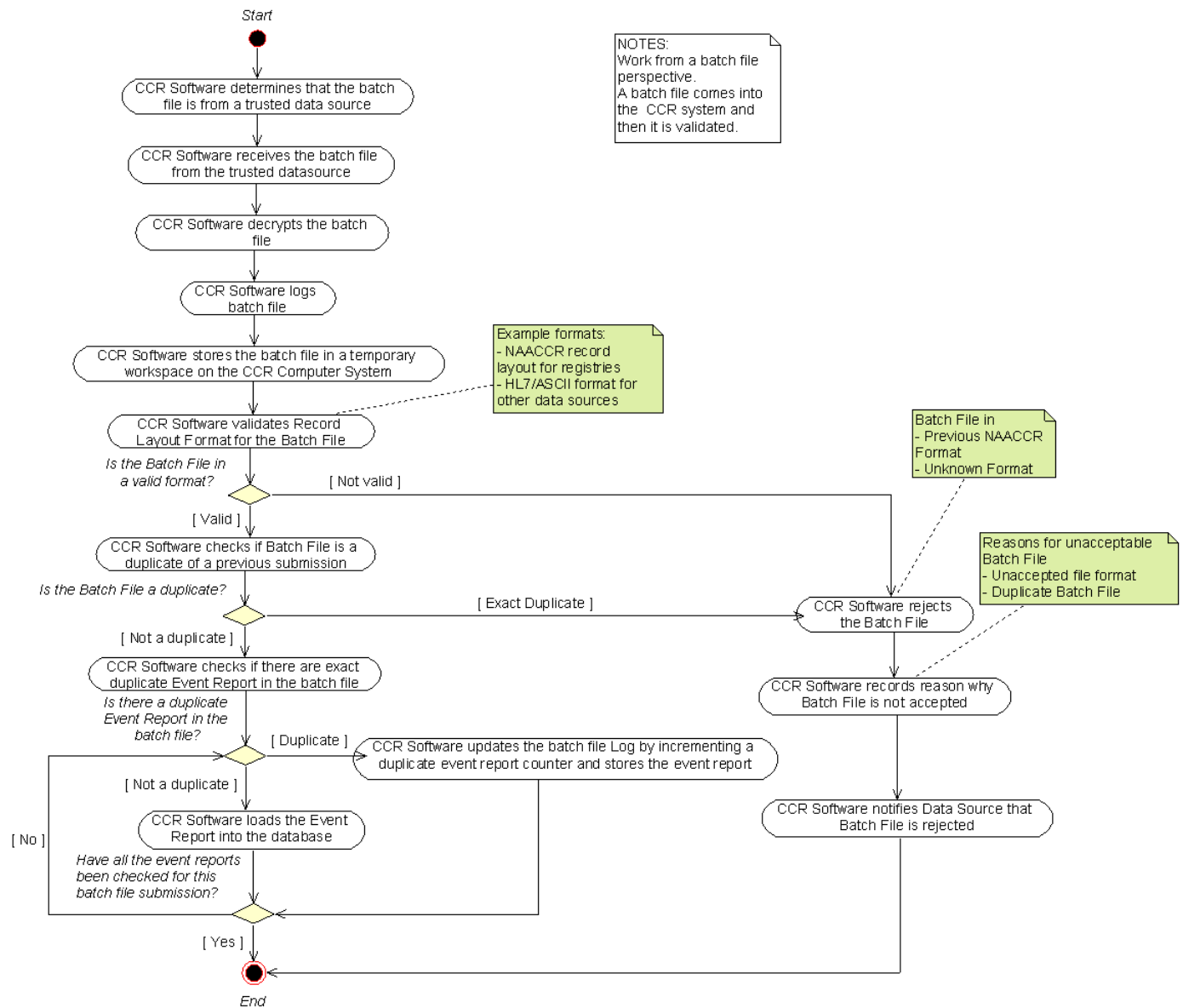None.

11.0 Assumptions

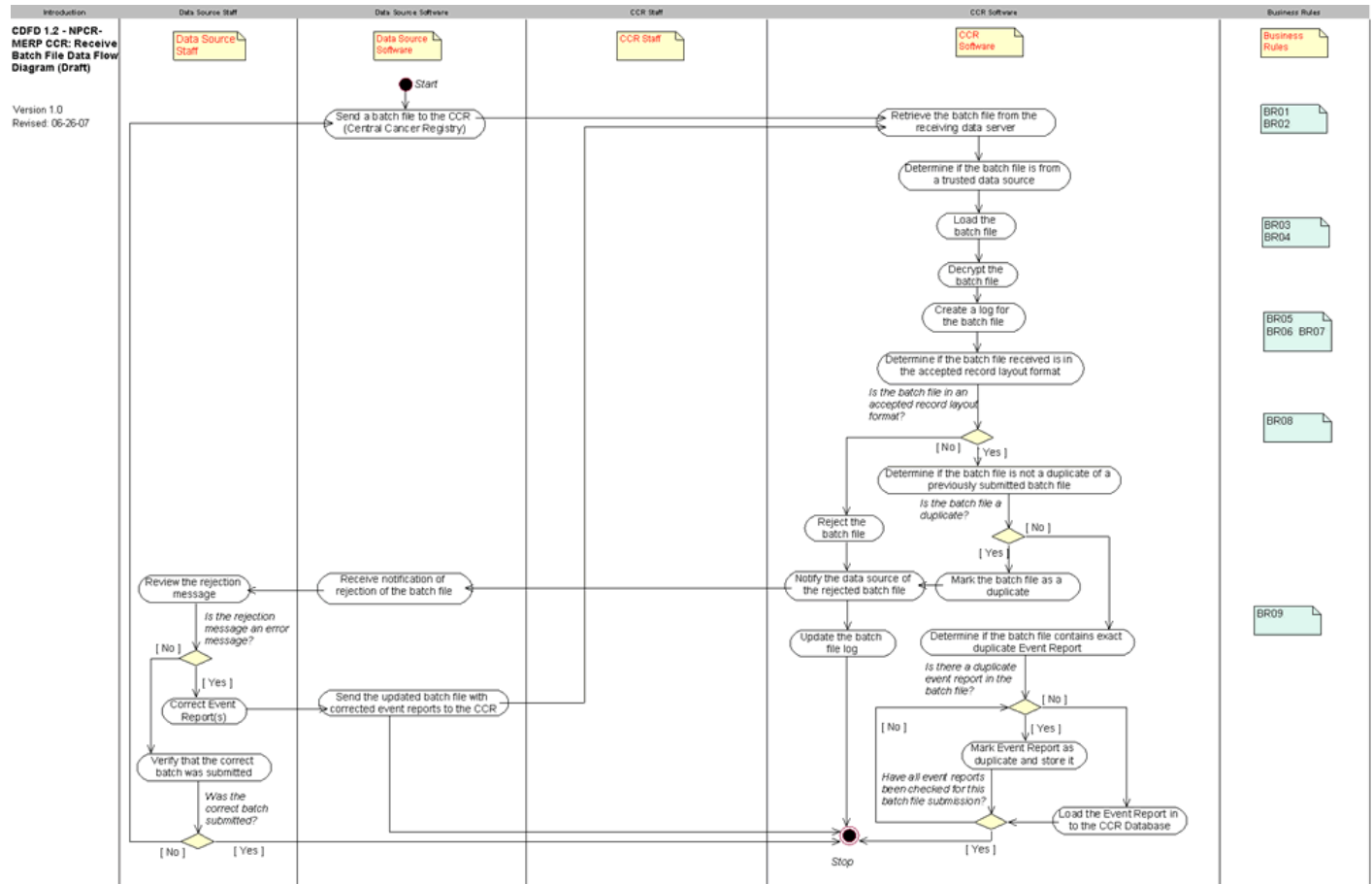Batch files are in electronic format.

12.0 Notes and Issues

None.

13.0 References

Baseline use case content provided in part by SEER*DMS design documents.

Appendix A: Receive Batch File Workflow Diagram

Appendix B: Receive Batch File Data Flow Diagram

Appendix C: Batch File Log Data Element Requirements

Data Element	Definition
CCR-assigned batch file number	Sequential number of batches received.
Date and time when the batch file was transmitted from the data source	
Date and time when the batch file was received in the central holding place (firewall transfer)	
Transmission status to the central holding place	Successful or failed.
Date and time when the batch file was transferred from the central holding place (firewall transfer)	
Date and time when the batch file was received by the CCR	Date when the files were received from the data source.
Transmission status to the CCR	Successful or failed.
Date when the batch file was loaded	Date when the files were loaded to the central Data Management System (DMS)
Sender or data source name	
File name	A batch can contain multiple files, and the files can be a combination of multiple types of source documents.
File signature	
E-mail address of sender	
Processing status	Pending, processing, or complete.
Reason for failure	
Type of records	A batch can contain multiple files, and the files can be a combination of multiple types of source documents.
Number of records	
Number of failed records	
Number of duplicate records	
Rollback eligibility	If a large number of records in a package are corrupt, users with the necessary privileges can roll back the package.
Rollback user ID	
Rollback date	
Rollback reason	

Appendix D: Naming Standards for Data Source Submissions

Purpose: To ensure accuracy and consistency of transmitting batch files to central registries.

A data source may be required to send data to more than one registry. To make this process as efficient as possible and to allow files to be identified and processed correctly by the recipient, a national consensus standard for naming batch files is recommended. Establishing a standard especially will assist data sources submitting to more than one central cancer registry.

The naming convention is DataSource.[ReportType].[SubmitType].DateOfTransmission.FileNumber

- **DataSource:** Required. A national provider ID or another unique identifier.

- **ReportType:** Optional. The type of report being submitted (see table 1).

- **SubmitType**: Optional. New, update, correction, or other.

- **DateOfTransmission:** Required. The date when the file was transmitted, in either CCYYMMDD.MM or CCYY[3-digit day of the year, 001–366] format.

- **FileNumber:** Required. The sequential number of the file, among the files submitted that day.

Table 1. Data Source Report Type Abbreviations

Abbreviation	Data Source Name	Comments
TR	Tumor or cancer registry abstract	
PATH	Pathology report	All pathology report types should use this report type abbreviation in the batch file name.
CLAIMS	Claims data	
MD_CLINIC	Physician or clinic office data	
DX_IMAGE	Diagnostic imaging	All imaging and radiology report types should use this report type abbreviation in the batch file name.
SURG	Surgery report	

Table 2. Submit Type Abbreviations

Abbreviation	Submit Type Name
NEW	New case
UPD	Update
FOL	Follow-up
DEL	Deletion

Appendix E: Using Signatures to Identify Duplicate Batch Files

California Cancer Registry

For validating uniqueness among files, the California Cancer Registry uses the TIdHashMessageDigest5 class, which is part of the Indy Project. The TIdHashMessageDigest5 class implements the RSA-MD5 message digest encryption algorithm.

The RSA-MD5 algorithm takes an input message of an arbitrary length and produces a 128-bit "fingerprint" or "message digest" of the input.

The input message supplied to the algorithm is a FileStream object containing the file being uploaded. In turn, the TIdHashMessageDigest5 class implements a method which converts this fingerprint to hexadecimal format, which is stored in the database. If the fingerprint of an incoming file matches a fingerprint (or,signature) already stored in the database, the entire file is marked as a duplicate and processing ends.

If the fingerprint of the incoming file does not match a fingerprint already stored in the database, the records in the file are processed one at a time.

NPCR's Web Plus

NPCR's Web Plus software uses the SHA1Managed class (available in the .NET framework) to produce the SHA1 hash (160-bit fingerprint) of files in Web Plus to check for duplicate files

Use Case Administrative Information

1. Use Case History

Version 1.0 presented to the NPCR–MERP Central Cancer Registry Workgroup January 8, 2008.

2. Created By

- NPCR–MERP Central Cancer Registry Workgroup
- NPCR–MERP Technical Development Team

3. Date Created

January 3, 2007

4. Last Updated By

MA

5. Date Last Updated

March 2, 2008

Revision History

Name	Date	Reason for Changes	Version
MA	1/3/07		0.01
MA	1/4/07	To finish the use case.	0.01
WKS, MA	2/1/07	Changed according to the Central Cancer Registry Workgroup discussion.	0.02
WKS, MA	4/9/07	Finalized business rules.	0.03
WKS, MA	4/19/07	Added purpose to the business rules table.	0.04
WKS, MA	4/19/07	Added discussion on standard file name convention.	0.05
WKS, MA	5/3/07	Added a new step to the Normal Course of Events.	0.05
MA	10/16/07	Edited the document.	0.06
MA	10/26/07	Integrated business rules.	0.07
WKS, MA	1/14/08	Finalized the use case.	1.0
MA	3/2/08	Updated the document after technical review.	1.0